# Appendix A

## Elements of Linear Algebra, Topology, and Calculus

### A.1 LINEAR ALGEBRA

We follow the usual conventions of matrix computations. $\mathbb{R}^{n\times p}$ is the set of all $n \times p$ real matrices ($m$ rows and $p$ columns). $\mathbb{R}^n$ is the set $\mathbb{R}^{n\times 1}$ of column vectors with $n$ real entries. $A(i,j)$ denotes the $i,j$ entry ($i$th row, $j$th column) of the matrix $A$. Given $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$, the matrix product $AB \in \mathbb{R}^{m\times p}$ is defined by $(AB)(i,j) = \sum_{k=1}^{n} A(i,k)B(k,j)$, $i = 1,\ldots,m$, $j = 1,\ldots,p$. $A^T$ is the *transpose* of the matrix $A$: $(A^T)(i,j) = A(j,i)$. The entries $A(i,i)$ form the *diagonal* of $A$. A matrix is *square* if is has the same number of rows and columns. When $A$ and $B$ are square matrices of the same dimension, $[A,B] = AB - BA$ is termed the *commutator* of $A$ and $B$. A matrix $A$ is *symmetric* if $A^T = A$ and *skew-symmetric* if $A^T = -A$. The commutator of two symmetric matrices or two skew-symmetric matrices is symmetric, and the commutator of a symmetric and a skew-symmetric matrix is skew-symmetric. The *trace* of $A$ is the sum of the diagonal elements of $A$,

$$\mathrm{tr}(A) = \sum_{i=1}^{\min(n,p)} A(i,i).$$

We have the following properties (assuming that $A$ and $B$ have adequate dimensions)

$$\mathrm{tr}(A) = \mathrm{tr}(A^T), \tag{A.1a}$$

$$\mathrm{tr}(AB) = \mathrm{tr}(BA), \tag{A.1b}$$

$$\mathrm{tr}([A,B]) = 0, \tag{A.1c}$$

$$\mathrm{tr}(B) = 0 \quad \text{if } B^T = -B, \tag{A.1d}$$

$$\mathrm{tr}(AB) = 0 \quad \text{if } A^T = A \text{ and } B^T = -B. \tag{A.1e}$$

An $n\times n$ matrix $A$ is *invertible* (or *nonsingular*) if there exists an $n\times n$ matrix $B$ such that $AB = BA = I_n$, where $I_n$ denotes the $n\times n$ *identity matrix* with ones on the diagonal and zeros everywhere else. If this is the case, then $B$ is uniquely determined by $A$ and is called the *inverse* of $A$, denoted by $A^{-1}$. A matrix that is not invertible is called *singular*. A matrix $Q$ is *orthonormal* if $Q^TQ = I$. A square orthonormal matrix is termed *orthogonal* and satisfies $Q^{-1} = Q^T$.

The notion of an $n$-dimensional *vector space* over $\mathbb{R}$ is an abstraction of $\mathbb{R}^n$ endowed with its operations of addition and multiplication by a scalar. Any

$n$-dimensional real vector space $\mathcal{E}$ is isomorphic to $\mathbb{R}^n$. However, producing a diffeomorphism involves generating a basis of $\mathcal{E}$, which may be computationally intractable; this is why most of the following material is presented on abstract vector spaces. We consider only finite-dimensional vector spaces over $\mathbb{R}$.

A *normed vector space* $\mathcal{E}$ is a vector space endowed with a *norm*, i.e., a mapping $x \in \mathcal{E} \mapsto \|x\| \in \mathbb{R}$ with the following properties. For all $a \in \mathbb{R}$ and all $x, y \in \mathcal{E}$,

1. $\|x\| \geq 0$, and $\|x\| = 0$ if and only if $x$ is the zero vector;
2. $\|ax\| = |a| \, \|x\|$ ;
3. $\|x + y\| \leq \|x\| + \|y\|$.

Given two normed vector spaces $\mathcal{E}$ and $\mathcal{F}$, a mapping $A : \mathcal{E} \mapsto \mathcal{F}$ is a *(linear) operator* if $A[\alpha x + \beta y] = \alpha A[x] + \beta A[y]$ for all $x, y \in \mathcal{E}$ and all $\alpha, \beta \in \mathbb{R}$. The set $\mathfrak{L}(\mathcal{E}; \mathcal{F})$ of all operators from $\mathcal{E}$ to $\mathcal{F}$ is a vector space. An operator $A \in \mathfrak{L}(\mathbb{R}^n; \mathbb{R}^m)$ can be represented by an $m \times n$ matrix (also denoted by $A$) such that $A[x] = Ax$ for all $x \in \mathbb{R}^n$. This representation is an isomorphism that matches the composition of operators with the multiplication of matrices. Let $\mathcal{E}$, $\mathcal{F}$, and $\mathcal{G}$ be normed vector spaces, let $\|\cdot\|_{\mathfrak{L}(\mathcal{E};\mathcal{F})}$ be a norm on $\mathfrak{L}(\mathcal{E}; \mathcal{F})$, $\|\cdot\|_{\mathfrak{L}(\mathcal{F};\mathcal{G})}$ be a norm on $\mathfrak{L}(\mathcal{F}; \mathcal{G})$, and $\|\cdot\|_{\mathfrak{L}(\mathcal{E};\mathcal{G})}$ be a norm on $\mathfrak{L}(\mathcal{E}; \mathcal{G})$. These norms are called *mutually consistent* if $\|B \circ A\|_{\mathfrak{L}(\mathcal{E};\mathcal{G})} \leq \|A\|_{\mathfrak{L}(\mathcal{E};\mathcal{F})} \|B\|_{\mathfrak{L}(\mathcal{F};\mathcal{G})}$ for all $A \in \mathfrak{L}(\mathcal{E}; \mathcal{F})$ and all $B \in \mathfrak{L}(\mathcal{F}; \mathcal{G})$. A *consistent* or *submultiplicative norm* is a norm that is mutually consistent with itself. The *operator norm* or *induced norm* of $A \in \mathfrak{L}(\mathcal{E}; \mathcal{F})$ is

$$\|A\| := \max_{x \in \mathcal{E}, x \neq 0} \frac{\|A[x]\|}{\|x\|}.$$

Operator norms are mutually consistent.

Given normed vector spaces $\mathcal{E}_1$, $\mathcal{E}_2$, and $\mathcal{F}$, a mapping $A$ from $\mathcal{E}_1 \times \mathcal{E}_2$ to $\mathcal{F}$ is called a *bilinear operator* if for any $x_2 \in \mathcal{E}_2$ the linear mapping $x_1 \mapsto A[x_1, x_2]$ is a linear operator from $\mathcal{E}_1$ to $\mathcal{F}$, and for any $x_1 \in \mathcal{E}_1$ the linear mapping $x_2 \mapsto A[x_1, x_2]$ is a linear operator from $\mathcal{E}_2$ to $\mathcal{F}$. The set of bilinear operators from $\mathcal{E}_1 \times \mathcal{E}_2$ to $\mathcal{F}$ is denoted by $\mathfrak{L}(\mathcal{E}_1, \mathcal{E}_2; \mathcal{F})$, and we use the notation $\mathfrak{L}_2(\mathcal{E}; \mathcal{F})$ for $\mathfrak{L}(\mathcal{E}, \mathcal{E}; \mathcal{F})$. These definitions are readily extended to multilinear operators. A bilinear operator $A \in \mathfrak{L}_2(\mathcal{E}; \mathcal{F})$ is *symmetric* if $A[x, y] = A[y, x]$ for all $x, y \in \mathcal{E}$. A symmetric bilinear operator $A \in \mathfrak{L}_2(\mathcal{E}, \mathbb{R})$ is *positive-definite* if $A[x, x] > 0$ for all $x \in \mathcal{E}$, $x \neq 0$.

By *Euclidean space* we mean a finite-dimensional vector space endowed with an inner product, i.e., a bilinear, symmetric positive-definite form $\langle \cdot, \cdot \rangle$. The canonical example is $\mathbb{R}^n$, endowed with the inner product

$$\langle x, y \rangle := x^T y.$$

An *orthonormal basis* of an $n$-dimensional Euclidean space $\mathcal{E}$ is a sequence $(e_1, \ldots, e_n)$ of elements of $\mathcal{E}$ such that

$$\langle e_i, e_j \rangle = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Given an orthonormal basis of $\mathcal{E}$, the mapping that sends the elements of $\mathcal{E}$ to their vectors of coordinates in $\mathbb{R}^d$ is an isomorphism (an invertible mapping that preserves the vector space structure and the inner product). An operator $T : \mathcal{E} \to \mathcal{E}$ is termed *symmetric* if $\langle T[x], y\rangle = \langle x, T[y]\rangle$ for all $x, y \in \mathcal{E}$. Given an operator $T : \mathcal{E} \to \mathcal{F}$ between two Euclidean spaces $\mathcal{E}$ and $\mathcal{F}$, the *adjoint* of $T$ is the operator $T^* : \mathcal{F} \to \mathcal{E}$ satisfying $\langle T[x], y\rangle = \langle x, T^*[y]\rangle$ for all $x \in \mathcal{E}$ and all $y \in \mathcal{F}$. The *kernel* of the operator $T$ is the linear subspace $\ker(T) = \{x \in \mathcal{E} : T[x] = 0\}$. The *range* (or *image*) of $T$ is the set $\mathrm{range}(T) = \{T[x] : x \in \mathcal{E}\}$. Given a linear subspace $\mathcal{S}$ of $\mathcal{E}$, the *orthogonal complement* of $\mathcal{S}$ is $\mathcal{S}^\perp = \{x \in \mathcal{E} : \langle x, y\rangle = 0 \text{ for all } y \in \mathcal{S}\}$. Given $x \in \mathcal{E}$, there is a unique decomposition $x = x_1 + x_2$ with $x_1 \in \mathcal{S}$ and $x_2 \in \mathcal{S}^\perp$; $x_1$ is the *orthogonal projection* of $x$ onto $\mathcal{S}$ and is denoted by $\Pi_{\mathcal{S}}(x)$. The *Moore-Penrose inverse* or *pseudo-inverse* of an operator $T$ is the operator

$$T^\dagger : \mathcal{F} \to \mathcal{E} : y \mapsto (T|_{(\ker(T))^\perp})^{-1}[\Pi_{\mathrm{range}(T)} y],$$

where the restriction $T|_{(\ker(T))^\perp} : (\ker(T))^\perp \to \mathrm{range}(T)$ is invertible by construction and $\Pi_{\mathrm{range}(T)}$ is the orthogonal projector in $\mathcal{F}$ onto $\mathrm{range}(T)$. The *Euclidean norm* on a Euclidean space $\mathcal{E}$ is

$$\|x\| := \sqrt{\langle x, x\rangle}.$$

The Euclidean norm on $\mathbb{R}^n$ is

$$\|x\| := \sqrt{x^T x}.$$

The Euclidean norm on $\mathbb{R}^{n\times p}$ endowed with the inner product $\langle X, Y\rangle = \mathrm{tr}(X^T Y)$ is the *Frobenius norm* given by $\|A\|_F = (\sum_{i,j}(A(i,j))^2)^{1/2}$. The operator norm on $\mathbb{R}^{n\times n} \simeq \mathfrak{L}(\mathbb{R}^n; \mathbb{R}^n)$, where $\mathbb{R}^n$ is endowed with its Euclidean norm, is the *spectral norm* given by

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)},$$

where $\lambda_{\max}(A^T A)$ is the largest eigenvalue of the positive-semidefinite matrix $A^T A$.

An operator $T$ in $\mathfrak{L}(\mathcal{E}; \mathcal{E})$ is *invertible* if for all $y \in \mathcal{E}$ there exists $x \in \mathcal{E}$ such that $y = T[x]$. An operator that is not invertible is termed *singular*. Let $\mathrm{id}_{\mathcal{E}}$ denote the identity operator on $\mathcal{E}$: $\mathrm{id}_{\mathcal{E}}[x] = x$ for all $x \in \mathcal{E}$. Let $T \in \mathfrak{L}(\mathcal{E}; \mathcal{E})$ be a symmetric operator. A real number $\lambda$ is an *eigenvalue* of $T$ if the operator $T - \lambda\,\mathrm{id}_{\mathcal{E}}$ is singular; any vector $x \neq 0$ in the kernel of $T - \lambda\,\mathrm{id}_{\mathcal{E}}$ is an *eigenvector* of $T$ corresponding to the eigenvalue $\lambda$.

References: [GVL96], [Hal74], [Die69].

## A.2 TOPOLOGY

A topology on a set $X$ is an abstraction of the notion of open sets in $\mathbb{R}^n$. Defining a topology on $X$ amounts to saying which subsets of $X$ are open while retaining certain properties satisfied by open sets in $\mathbb{R}^n$. Specifically, a *topology* on a set $X$ is a collection $\mathcal{T}$ of subsets of $X$, called *open sets*, such that

1. $X$ and $\emptyset$ belong to $\mathcal{T}$;
2. the union of the elements of any subcollection of $\mathcal{T}$ is in $\mathcal{T}$;
3. the intersection of the elements of any finite subcollection of $\mathcal{T}$ is in $\mathcal{T}$.

A *topological space* is a couple $(X, \mathcal{T})$ where $X$ is a set and $\mathcal{T}$ is a topology on $X$. When the topology is clear from the context or is irrelevant, we simply refer to the topological space $X$.

Let $X$ be a topological space. A subset $A$ of $X$ is said to be *closed* if the set $X - A := \{x \in X : x \notin A\}$ is open. A *neighborhood* of a point $x \in X$ is a subset of $X$ that includes an open set containing $x$. A *limit point* (or *accumulation point*) of a subset $A$ of $X$ is a point $x$ of $X$ such that every neighborhood of $x$ intersects $A$ in some point other than $x$ itself. A subset of $X$ is closed if and only if it contains all its limit points. A sequence $\{x_k\}_{i=1,2,\ldots}$ of points of $X$ *converges* to the point $x \in X$ if, for every neighborhood $U$ of $x$, there is a positive integer $K$ such that $x_k$ belongs to $U$ for all $k \geq K$.

In view of the limited number of axioms that a topology has to satisfy, it is not suprising that certain properties that hold in $\mathbb{R}^n$ do not hold for an arbitrary topology. For example, singletons (subsets containing only one element) may not be closed; this is the case for the overlapping interval topology, a topology of $[-1, 1]$ whose open sets are intervals of the form $[-1, b)$ for $b > 0$, $(a, 1]$ for $a < 0$, and $(a, b)$ for $a < 0$, $b > 0$. Another example is that sequences may converge to more than one point; this is the case with the cofinite topology of an infinite set, whose open sets are all the subsets whose complements are finite (i.e., have finitely many elements). To avoid these strange situations, the following *separation axioms* have been introduced.

Let $X$ be a topological space. $X$ is $T_1$, or *accessible* or *Fréchet*, if for any distinct points $x$ and $y$ of $X$, there is an open set that contains $x$ and not $y$. Equivalently, every singleton is closed. $X$ is $T_2$, or *Hausdorff*, if any two distinct points of $X$ have disjoint neighborhoods. If $X$ is Hausdorff, then every sequence of points of $X$ converges to at most one point of $X$.

Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two topologies on the same set $X$. If $\mathcal{T}_1 \subseteq \mathcal{T}_2$, we say that $\mathcal{T}_2$ is *finer* than $\mathcal{T}_1$.

A *basis* for a topology on a set $X$ is a collection $\mathcal{B}$ of subsets of $X$ such that

1. each $x \in X$ belongs to at least one element of $\mathcal{B}$;
2. if $x \in (B_1 \cap B_2)$ with $B_1, B_2 \in \mathcal{B}$, then there exists $B_3 \in \mathcal{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2$.

If $\mathcal{B}$ is a basis for a topology $\mathcal{T}$ on $X$, then $\mathcal{T}$ equals the collection of all unions of elements of $\mathcal{B}$. A topological space $X$ is called *second-countable* if it has a countable basis (i.e., a basis with countably many elements) for its topology.

Let $X$ and $Y$ be topological spaces. The *product topology* on $X \times Y$ is the topology having as a basis the collection $\mathcal{B}$ of all sets of the form $U \times V$, where $U$ is an open subset of $X$ and $V$ is an open subset of $Y$.

If $Y$ is a subset of a topological space $(X, \mathcal{T})$, then the collection $\mathcal{T}_Y = \{Y \cap U : U \in \mathcal{T}\}$ is a topology on $Y$ called the *subspace topology*.

If $\sim$ is an equivalence relation on a topological space $X$, then the collection of all subsets $U$ of the quotient set $X/\sim$ such that $\pi^{-1}(U)$ is open in $X$ is called the *quotient topology* of $X/\sim$. (We refer the reader to Section 3.4 for a discussion of the notions of equivalence relation and quotient set.)

Subspaces and products of Hausdorff spaces are Hausdorff, but quotient spaces of Hausdorff spaces need not be Hausdorff. Subspaces and countable products of second-countable spaces are second-countable, but quotients of second-countable spaces need not be second-countable.

Let $X$ be a topological space. A collection $\mathcal{A}$ of subsets of $X$ is said to *cover* $X$, or to be a *covering* of $X$, if the union of the elements of $\mathcal{A}$ is equal to $X$. It is called an open covering of $X$ if its elements are open subsets of $X$. The space $X$ is said to be *compact* if every open covering $\mathcal{A}$ of $X$ contains a finite subcollection that also covers $X$. The *Heine-Borel theorem* states that a subset of $\mathbb{R}^n$ (with the subspace topology) is compact if and only if it is closed and bounded.

Let $\mathbb{F}$ denote either $\mathbb{R}$ or $\mathbb{C}$. The set $\mathbb{F}^n$ has a standard topology; the collection of "open balls" $\{y \in \mathbb{F}^n : \sum_i |y_i - x_i|^2 < \epsilon\}$, $x \in \mathbb{F}^n$, $\epsilon > 0$, is a basis of that topology. A finite-dimensional vector space $\mathcal{E}$ over a field $\mathbb{F}$ ($\mathbb{R}$ or $\mathbb{C}$) inherits a natural topology: let $F : \mathcal{E} \to \mathbb{F}^n$ be an isomorphism of $\mathcal{E}$ with $\mathbb{F}^n$ and endow $\mathcal{E}$ with the topology where a subset $X$ of $\mathcal{E}$ is open if and only if $F(X)$ is open in $\mathbb{F}^n$. Hence, $\mathbb{R}^{n\times p}$ has a natural topology as a finite-dimensional vector space, and the noncompact Stiefel manifold $\mathbb{R}_*^{n\times p}$ has a natural topology as a subset of $\mathbb{R}^{n\times p}$.

Reference: [Mun00].

## A.3 FUNCTIONS

There is no general agreement on the way to define a function, its range, and its domain, so we find it useful to state our conventions. A *function* (or *map*, or *mapping*)

$$f : A \to B$$

is a set of ordered pairs $(a, b)$, $a \in A$, $b \in B$, with the property that, if $(a, b)$ and $(a, c)$ are in the set, then $b = c$. If $(a, b) \in f$, we write $b$ as $f(a)$. Note that we do not require that $f(a)$ be defined for all $a \in A$. This is convenient, as it allows us to simply say, for example, that the tangent is a function from $\mathbb{R}$ to $\mathbb{R}$. The *domain* of $f$ is $\mathrm{dom}(f) := \{a \in A : \exists b \in B : (a, b) \in f\}$, and the *range* (or *image*) of $f$ is $\mathrm{range}(f) := \{b \in B : \exists a \in A : (a, b) \in f\}$. If $\mathrm{dom}(f) = A$, then $f$ is said to be *on* $A$. If $\mathrm{range}(f) = B$, then $f$ is *onto* $B$. An onto function is also called a *surjection*. An *injection* is a function $f$ with the property that if $x \neq y$, then $f(x) \neq f(y)$. A function from a set $A$ to a set $B$ is a *bijection* or a *one-to-one correspondence* if it is both an injection and a surjection from $A$ to $B$. The *preimage* $f^{-1}(Y)$ of a set $Y \subseteq B$ under

$f$ is the subset of $A$ defined by

$$f^{-1}(Y) = \{x \in A : f(x) \in Y\}.$$

Given $y \in B$, the set $f^{-1}(y) := f^{-1}(\{y\})$ is called a *fiber* or *level set* of $f$.

A function $f : A \to B$ between two topological spaces is said to be *continuous* if for each open subset $V$ of $B$, the set $f^{-1}(V)$ is an open subset of $A$. By the *extreme value theorem*, if a real-valued function $f$ is continuous on a compact set $X$, then there exist points $c$ and $d$ in $X$ such that $f(c) \leq f(x) \leq f(d)$ for every $x \in X$.

## A.4 ASYMPTOTIC NOTATION

Let $\mathcal{E}, \mathcal{F}, \mathcal{G}$ be normed vector spaces and let $F : \mathcal{E} \to \mathcal{F}$ and $G : \mathcal{E} \to \mathcal{G}$ be defined on a neighborhood of $x_* \in \mathcal{E} \cup \{\infty\}$. The notation

$$F(x) = O(G(x)) \text{ as } x \to x_*$$

(or simply $F(x) = O(G(x))$ when $x_*$ is clear from the context) means that

$$\limsup_{\substack{x \to x_* \\ x \neq x_*}} \frac{\|F(x)\|}{\|G(x)\|} < \infty.$$

In other words, $F(x) = O(G(x))$ as $x \to x_*$, $x_* \in \mathcal{E}$, means that there is $C \geq 0$ and $\delta > 0$ such that

$$\|F(x)\| \leq C\|G(x)\| \tag{A.2}$$

for all $x$ with $\|x - x_*\| < \delta$, and $F(x) = O(G(x))$ as $x \to \infty$ means that there is $C \geq 0$ and $\delta > 0$ such that (A.2) holds for all $x$ with $\|x\| > \delta$. The notation

$$F(x) = o(G(x)) \text{ as } x \to x_*$$

means that

$$\lim_{\substack{x \to x_* \\ x \neq x_*}} \frac{\|F(x)\|}{\|G(x)\|} = 0.$$

Finally, the notation

$$F(x) = \Omega(G(x)) \text{ as } x \to x_*$$

means that there exist $C > 0$, $c > 0$, and a neighborhood $\mathcal{N}$ of $x_*$ such that

$$c\|G(x)\| \leq \|F(x)\| \leq C\|G(x)\|$$

for all $x \in \mathcal{N}$.

We use similar notation to compare two sequences $\{x_k\}$ and $\{y_k\}$ in two normed spaces. The notation $y_k = O(x_k)$ means that there is $C \geq 0$ such that

$$\|y_k\| \leq C\|x_k\|$$

for all $k$ sufficiently large. The notation $y_k = o(x_k)$ means that

$$\lim_{k \to \infty} \frac{\|y_k\|}{\|x_k\|} = 0.$$

The notation $y_k = \Omega(x_k)$ means that there exist $C > 0$ and $c > 0$ such that

$$c\|x_k\| \leq \|y_k\| \leq C\|x_k\|$$

for all $k$ sufficiently large.

The loose notation $y_k = O(x_k^p)$ is used to denote that $\|y_k\| = O(\|x_k\|^p)$, and likewise for $o$ and $\Omega$.

## A.5 DERIVATIVES

We present the concept of a derivative for functions between two finite-dimensional normed vector spaces. The extension to manifolds can be found in Chapter 3.

Let $\mathcal{E}$ and $\mathcal{F}$ be two finite-dimensional vector spaces over $\mathbb{R}$. (A particular case is $\mathcal{E} = \mathbb{R}^m$ and $\mathcal{F} = \mathbb{R}^n$.) A function $F : \mathcal{E} \to \mathcal{F}$ is *(Fréchet)-differentiable* at a point $x \in \mathcal{E}$ if there exists a linear operator

$$\mathrm{D}F(x) : \mathcal{E} \to \mathcal{F} : h \mapsto \mathrm{D}F(x)[h],$$

called the *(Fréchet) differential* (or the *Fréchet derivative*) of $F$ at $x$, such that

$$F(x + h) = F(x) + \mathrm{D}F(x)[h] + o(\|h\|);$$

in other words,

$$\lim_{y \to x} \frac{\|F(y) - F(x) - \mathrm{D}F(x)[y - x]\|}{\|y - x\|} = 0.$$

The element $\mathrm{D}F(x)[h] \in \mathcal{F}$ is called the *directional derivative* of $F$ at $x$ along $h$. (We use the same notation $\mathrm{D}F(x)$ for the differential of a function $F$ between two manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$; then $\mathrm{D}F(x)$ is a linear operator from the vector space $T_x\mathcal{M}_1$ to the vector space $T_{F(x)}\mathcal{M}_2$; see Section 3.5.)

By convention, the notation "D" applies to the expression that follows. Hence $\mathrm{D}(f \circ g)(x)$ and $\mathrm{D}f(g(x))$ are two different things: the derivative of $f \circ g$ at $x$ for the former, the derivative of $f$ at $g(x)$ for the latter. We have the *chain rule*

$$\mathrm{D}(f \circ g)(x) = \mathrm{D}f(g(x)) \circ \mathrm{D}g(x);$$

i.e.,

$$\mathrm{D}(f \circ g)(x)[h] = \mathrm{D}f(g(x))\,[\mathrm{D}g(x)[h]]$$

for all $h$.

The function $F : \mathcal{E} \to \mathcal{F}$ is said to be *differentiable on an open domain* $\Omega \subseteq \mathcal{E}$ if $F$ is differentiable at every point $x \in \Omega$. Note that, for all $x \in \Omega$,

$\mathrm{D}F(x)$ belongs to the vector space $\mathfrak{L}(\mathcal{E};\mathcal{F})$ of all linear operators from $\mathcal{E}$ to $\mathcal{F}$, which is itself a normed vector space with the induced norm

$$\|u\| = \max_{\|x\|=1} \|u(x)\|, \quad u \in \mathfrak{L}(\mathcal{E};\mathcal{F}).$$

The function $F$ is termed *continuously differentiable* (or $C^1$) on the open domain $\Omega$ if $\mathrm{D}F : \mathcal{E} \to \mathfrak{L}(\mathcal{E};\mathcal{F})$ is continuous on $\Omega$.

Assume that bases $(e_1, \ldots, e_m)$ and $(e'_1, \ldots, e'_n)$ are given for $\mathcal{E}$ and $\mathcal{F}$ and let $\hat{F} : \mathbb{R}^m \to \mathbb{R}^n$ be the expression of $F$ in these bases; i.e., $\sum_j \hat{F}^j(\hat{x})e'_j = F(\sum_i \hat{x}^i e_i)$. Then $F$ is continuously differentiable if and only if the partial derivatives of $\hat{F}$ exist and are continuous, and we have

$$\mathrm{D}F(x)[h] = \sum_j \sum_i \partial_i \hat{F}^j(\hat{x})\hat{h}^i e'_j.$$

It can be shown that this expression does not depend on the chosen bases.

If $f$ is a real-valued function on a Euclidean space $\mathcal{E}$, then, given $x \in \mathcal{E}$, we define $\operatorname{grad} f(x)$, the *gradient* of $f$ at $x$, as the unique element of $\mathcal{E}$ that satisfies

$$\langle \operatorname{grad} f(x), h\rangle = \mathrm{D}f(x)[h] \quad \text{for all } h \in \mathcal{E}.$$

Given an orthonormal basis $(e_1, \ldots, e_d)$ of $\mathcal{E}$, we have

$$\operatorname{grad} f(x) = \sum_i \partial_i \widehat{f}(x^1, \ldots, x^d)e_i,$$

where $x^1 e_1 + \cdots + x^d e_d = x$ and $\widehat{f}$ is the expression of $f$ in the basis.

If $F : \mathcal{E} \to \mathcal{F}$ is a linear function, then $\mathrm{D}F(x)[h] = F(h)$ for all $x, h \in \mathcal{E}$. In particular, the derivative of the function $\operatorname{tr} : \mathbb{R}^{n\times n} \to \mathbb{R}$ is given by

$$\mathrm{D}\operatorname{tr}(X)[H] = \operatorname{tr}(H).$$

For the function $\operatorname{inv} : \mathbb{R}^{p\times p}_* \to \mathbb{R}^{p\times p}_* : M \mapsto M^{-1}$, we have

$$\mathrm{D}\operatorname{inv}(X)[Z] = -X^{-1}ZX^{-1}.$$

In other words, if $t \mapsto X(t)$ is a smooth curve in the set of invertible matrices, then

$$\frac{\mathrm{d}X^{-1}}{\mathrm{d}t} = -X^{-1}\frac{\mathrm{d}X}{\mathrm{d}t}X^{-1}. \tag{A.3}$$

The derivative of the determinant is given by *Jacobi's formula*,

$$\mathrm{D}\det(X)[Z] = \operatorname{tr}(\operatorname{adj}(X)Z),$$

where $\operatorname{adj}(X) := \det(X)X^{-1}$. For $X \in \mathbb{R}^{n\times p}$, let $\operatorname{qf}(X)$ denote the $Q$ factor of the *thin QR decomposition* $X = QR$, where $Q \in \mathbb{R}^{n\times p}$ is orthonormal and $R \in \mathbb{R}^{p\times p}$ is upper triangular with strictly positive diagonal elements. We have

$$\mathrm{D}\operatorname{qf}(X)[Z] = X\rho_{\text{skew}}(Q^TZR^{-1}) + (I - XX^T)ZR^{-1},$$

where $X = QR$ is the thin $QR$ decomposition of $X$ and $\rho_{\text{skew}}(A)$ denotes the skew-symmetric part of the decomposition of $A$ into the sum of a skew-symmetric matrix and an upper triangular matrix.

If the mapping $\mathrm{D}F : \mathcal{E} \to \mathfrak{L}(\mathcal{E};\mathcal{F})$ is differentiable at a point $x \in \mathcal{E}$, we say that $F$ is *twice differentiable* at $x$. The differential of $\mathrm{D}F$ at $x$ is called the *second derivative* of $F$ at $x$ and denoted by $\mathrm{D}^2F(x)$. This is an element of $\mathfrak{L}(\mathcal{E};\mathfrak{L}(\mathcal{E};\mathcal{F}))$, but this space is naturally identified with the space $\mathfrak{L}_2(\mathcal{E};\mathcal{F})$ of bilinear mappings of $\mathcal{E}\times\mathcal{E}$ into $\mathcal{F}$, and we use the notation $\mathrm{D}^2F(x)[h_1,h_2]$ for $(\mathrm{D}^2F(x)[h_1])[h_2]$. The second derivative satisfies the symmetry property $\mathrm{D}^2F(x)[h_1,h_2] = \mathrm{D}^2F(x)[h_2,h_1]$ for all $x, h_1, h_2 \in \mathcal{E}$. If $g \in \mathcal{E}$ and $h$ is a differentiable function on $\mathcal{E}$ into $\mathcal{E}$, then

$$\mathrm{D}\left(\mathrm{D}F\left(\cdot\right)\left[h(\cdot)\right]\right)(x)\left[g\right] = \mathrm{D}^2F(x)[h(x),g] + \mathrm{D}F\left(x\right)\left[\mathrm{D}h\left(x\right)\left[g\right]\right].$$

If $\mathcal{E}$ is a Euclidean space and $f$ is a twice-differentiable, real-valued function on $\mathcal{E}$, then the unique symmetric operator $\operatorname{Hess} f(x) : \mathcal{E} \to \mathcal{E}$ defined by

$$\langle \operatorname{Hess} f(x)[h_1], h_2\rangle = \mathrm{D}^2 f(x)[h_1,h_2] \quad \text{for all } h_1,h_2 \in \mathcal{E},$$

is termed the *Hessian operator* of $f$ at $x$. We have

$$\operatorname{Hess} f(x)[h] = \mathrm{D}(\operatorname{grad} f)(x)[h]$$

for all $h \in \mathcal{E}$. Given an orthonormal basis $(e_1,\ldots,e_d)$ of $\mathcal{E}$, we have

$$\operatorname{Hess} f(x)[e_i] = \sum_j \partial_i\partial_j \widehat{f}(x^1,\ldots,x^d)e_j$$

and

$$\mathrm{D}^2 f(x)[e_i,e_j] = \partial_i\partial_j\widehat{f}(x^1,\ldots,x^d),$$

where $\widehat{f}$ is the expression of $f$ in the basis.

The definition of the second derivative is readily generalized to derivatives of higher order. By induction on $p$, we define a *$p$-times-differentiable mapping* $F : \mathcal{E} \to \mathcal{F}$ as a $(p-1)$-times-differentiable mapping whose $(p-1)$th derivative $\mathrm{D}^{p-1}F$ is differentiable, and we call the derivative $\mathrm{D}(\mathrm{D}^{p-1}F)$ the *$p$th derivative* of $F$, written $\mathrm{D}^pF$. The element $\mathrm{D}^pF(x)$ is identified with an element of the space $\mathfrak{L}_p(\mathcal{E};\mathcal{F})$ of the $p$-linear mappings of $\mathcal{E}$ into $\mathcal{F}$, and we write it

$$(h_1,\ldots,h_p) \mapsto \mathrm{D}^pF(x)[h_1,\ldots,h_p].$$

The function $F : \mathbb{R}^n \to \mathbb{R}^m$ is *smooth* if it is continuously differentiable to all orders. This happens if and only if the partial derivatives of $\hat{F}$ exist and are continuous to all orders, and we have

$$\mathrm{D}^pF(x)[h_1,\ldots,h_p] = \sum_j \sum_{i_1,\ldots,i_p} \partial_{i_1}\cdots\partial_{i_p}\hat{F}^j(\hat{x})\hat{h}_1^{i_1}\cdots\hat{h}_p^{i_p}\ e'_j,$$

where the superscripts of $\hat{h}_1,\ldots,\hat{h}_p$ denote component numbers.

Reference: [Die69] and [Deh95] for the derivatives of several matrix functions.

## A.6 TAYLOR'S FORMULA

Let $\mathcal{E}$ and $\mathcal{F}$ be two finite-dimensional normed vector spaces and let $F : \mathcal{E} \to \mathcal{F}$ be $(p+1)$-times continuously differentiable on an open convex domain $\Omega \subseteq \mathcal{E}$. (The set $\Omega$ is *convex* if it contains all the line segments connecting any pair of its points.) *Taylor's theorem* (with the remainder in *Cauchy form*) states that, for all $x$ and $x+h$ in $\Omega$,

$$F(x+h) = F(x) + \frac{1}{1!}\mathrm{D}F(x)[h] + \frac{1}{2!}\mathrm{D}^2F(x)[h,h] + \cdots + \frac{1}{p!}\mathrm{D}^pF(x)[h,\dots,h] + R_p(h;x), \quad \text{(A.4)}$$

where

$$R_p(h;x) = \int_0^1 \frac{(1-t)^p}{p!}\mathrm{D}^{p+1}F(x+th)[h,\dots,h]\,\mathrm{d}t = O(\|h\|^{p+1}).$$

If $F$ is real-valued, then the remainder $R_p(h;x)$ can also be expressed in *Lagrange form*: for all $x$ and $x+h$ in $\Omega$, there exists $t \in (0,1)$ such that

$$F(x+h) = F(x) + \frac{1}{1!}\mathrm{D}F(x)[h] + \frac{1}{2!}\mathrm{D}^2F(x)[h,h] + \cdots + \frac{1}{p!}\mathrm{D}^pF(x)[h,\dots,h] + \frac{1}{(p+1)!}\mathrm{D}^{p+1}F(x+th)[h,\dots,h]. \quad \text{(A.5)}$$

The function

$$h \mapsto F(x) + \frac{1}{1!}\mathrm{D}F(x)[h] + \frac{1}{2!}\mathrm{D}^2F(x)[h,h] + \cdots + \frac{1}{p!}\mathrm{D}^pF(x)[h,\dots,h]$$

is called the $p$th-order *Taylor expansion* of $F$ around $x$.

The result $R_p(h;x) = O(\|h\|^{p+1})$ can be obtained under a weaker differentiability assumption. A function $G$ between two normed vector spaces $\mathcal{A}$ and $\mathcal{B}$ is said to be *Lipschitz-continuous* at $x \in \mathcal{A}$ if there exist an open set $\mathcal{U} \subseteq \mathcal{A}$, $x \in \mathcal{U}$, and a constant $\alpha$ such that for all $y \in \mathcal{U}$,

$$\|G(y) - G(x)\| \leq \alpha\|y - x\|. \quad \text{(A.6)}$$

The constant $\alpha$ is called a *Lipschitz constant* for $G$ at $x$. If this holds for a specific $\mathcal{U}$, then $G$ is said to be Lipschitz-continuous at $x$ in the neighborhood $\mathcal{U}$. If (A.6) holds for every $x \in \mathcal{U}$, then $G$ is said to be Lipschitz-continuous in $\mathcal{U}$ with Lipschitz constant $\alpha$. If $G$ is continuously differentiable, then it is Lipschitz-continuous in any bounded domain $\mathcal{U}$.

**Proposition A.6.1** *Let $\mathcal{E}$ and $\mathcal{F}$ be two finite-dimensional normed vector spaces, let $F : \mathcal{E} \to \mathcal{F}$ be $p$-times continuously differentiable in an open convex set $\mathcal{U} \subseteq \mathcal{E}$, $x \in \mathcal{U}$, and let the differential $\mathrm{D}^pF : \mathcal{E} \to \mathfrak{L}_p(\mathcal{E};\mathcal{F})$ be Lipschitz continuous at $x$ in the neighborhood $\mathcal{U}$ with Lipschitz constant $\alpha$ (using the induced norm in $\mathfrak{L}_p(\mathcal{E};\mathcal{F})$). Then, for any $x+h \in \mathcal{U}$,*

$$\left\| F(x+h) - F(x) - \frac{1}{1!}\mathrm{D}F(x)[h] - \cdots - \frac{1}{p!}\mathrm{D}^pF(x)[h,\dots,h] \right\| \leq \frac{\alpha}{(p+1)!}\|h\|^{p+1}.$$

*In particular, for* $p = 1$*, i.e.,* $F$ *continuously differentiable with a Lipschitz-continuous differential, we have*

$$\|F(x+h) - F(x) - \mathrm{D}F\,(x)\,[h]\,\| \leq \frac{\alpha}{2}\|h\|^2.$$